Flare Network Interface to Navisphere

Dean Throop

October 6, 1999

1 Introduction

This proposes a revised interface to Flare to allow the Navisphere Agent to manage Flare over TCP/IP.

The original Alpine Phase II Ethernet Service Port Functional Specificati on proposed using the existing serial line protocol over a TCP connection. However upon further consideration this approach was found unappealing because of synchronization problems with the current serial line protocol and the number of serial line specific manipulation calls made by the Agent.

A better interface would be to pass SCSI requests over TCP. This would standardize the management interface on the SCSI CDBs already defined and treat TCP/IP as an alternative transport to the Fibre Channel interface.

This means it will be possible to issue arbitrary SCSI requests such as read or write over the network. In general read or write requests will not be issued over the network as Fibre is the optimal interface for that. However there doesn't seem to be any reason to prohibit doing exactly that. It might be a useful to be able to generate read or write requests for testing. It might also be nice to have a back door to access LUN data for test verification or backup. It doesn't make sense to use it as the primary data storage data path but the ability will be there for completeness.

2 Overview

The goal of this new interface is to treat each TCP connection from a Navisphere Agent to Flare as a separate SCSI initiator. The Navisphere Agent can send multiple SCSI requests sent over the TCP connection and those requests will be processed according to SCSI rules for reordering and parallelism. Replies can come back in any order.

Completely modeling SCSI over Fibre on TCP would create some inefficiencies so the interface will be somewhat modified. To completely model SCSI over Fibre, the Agent would send a request to Flare and then Flare would request transfer of the data buffer associated with the request. This would mean an extra round trip message for Flare to read a data buffer. Rather than having Flare send requests to retrieve a data buffer from the Agent, the Agent will always send the data buffers that will be read by Flare as part of the request. For data buffers that are written by Flare, Flare will return these to the Agent as soon as processing has completed the buffer. When the Agent sends a request, TCP flow control may prevent the Agent write operation from completing immediately. If the Agent has several requests outstanding at one time, it should always keep accepting data (have a separate thread always doing a read or equivalent) so Flare can return completed requests to free resources to process incoming requests.

Having the Navisphere Agent send the data buffer with requests in which Flare will read the data buffer means the buffer will be transferred unnecessarily if the SCSI request has an error. Errors are expected to be infrequent and the saving of an extra round trip message should more than offset this.

3 References

[1] CLARiiON Disk-Array Licensed Internal Code Specification, Full-Fibre Storage Systems, Drawing No: 009-001854-04

[2] Storage Centric Setup Command Interface Rev 1.3
Charlie Hopkins 5/7/99

4 Environment

4.1 Security

Flare will only service connections from the Navisphere agent if the IP address of the agent matches one of a configured set of addresses that Flare will trust. The Flare administrator must configure the list of addresses as part of Flare management. Each address will have an associated mask to allow the administrator the ability to match an entire subnet with a single entry.

The Flare network interface only offers a modest level of security. Administrators can telnet to Flare where access is granted to anyone presenting the proper password. The password will be passed over the network in clear text so anyone on the network with a sniffer can obtain the password. If someone obtains the Flare password, they can access Flare and unbind LUNs and zero disks thus destroying data. Sites that are concerned about security should only connect Flare to a physically secure LAN segment trusted for management use and protected from other access (the site should use a firewall or take equivalent precautions to limit network connectivity).

Because the address from which Flare will accept connections must be specified in advance, the hosts running the Navisphere Agent must use fixed IP address; they can not use DHCP or some other dynamic mechanism of obtaining an address. This limitation results from the use of static addresses to grant access. If this proves to be a serious problem, we'll consider adding a more flexible mechanism in the future.

4.2 Network Performance

If Flare is connected to a network with a lot of broadcast traffic, that traffic could cause a performance impact on Flare. Flare should only be connected to a network where the level of broadcast traffic is known to be low enough to not adversely impact the performance of hosts on that network.

4.3 Encoding

The SCSI requests will be encoded using CTLD as described in the Storage Centric spec[2]. SCSI requests will be encoded with tags. The fields will be structured similarly to fibre requests and will include a Request Id to allow the Agent to match replies with requests. Data buffers transferred from Flare to the Agent will be returned independent of, and may proceed the status of the request that generated the data.

The CTLD encoding must always use the Binary encoding. All CDBs and data buffers will be in network byte order just as they would be constructed for the Fibre interface. Request Ids must be 4 bytes long.

5 Messages

5.1 SCSI Request

This message sends a SCSI request from the Agent to Flare.

Tag = TAG_NET_SCSI_REQUEST
Sub Tag = TAG_NET_REQUEST_ID
Sub Tag = TAG_NET_SCSI_CMND_PAYLOAD
Sub Tag = TAG_NET_BUFFER

The TAG_NET_REQUEST_ID is uninterpreted by FLARE and will be returned in the reply message.

The TAG_NET_SCSI_CMND_PAYLOAD contains a FCP CMND Payload as described in the CLARIION Fibre Spec[1].

The TAG_NET_BUFFER is a conditional Sub Tag; if Flare will read the data buffer from the Agent for the CDB, this provides that data. If Flare will return a data buffer to the Agent, this Sub Tag should not be present.

5.2 SCSI Reply

This message returns the status of a SCSI request from Flare to the Agent.

Tag = TAG_NET_SCSI_REPLY
Sub Tag = TAG_NET_REQUEST_ID
Sub Tag = TAG_NET_SCSI_RSP_PAYLOAD

The TAG_NET_REQUEST_ID returns the Request Id of the request.

The TAG_NET_SCSI_RSP_PAYLOAD returns a FCP RSP Payload as described in the Clariion Fibre spec[1].

5.3 SCSI Buffer

This message returns the data generated by Flare when processing a SCSI request. It may proceed or follow the SCSI status for the request.

Tag = TAG_NET_SCSI_RESULT_BUFFER
Sub Tag = TAG_NET_REQUET_ID
Sub Tag = TAG_NET_SCSI_BUFFER

The TAG_NET_REQUEST_ID returns the Request Id of the request that generated this reply buffer.

The TAG_NET_SCSI_BUFFER returns the data buffer generated by Flare.

5.4 Option Request

This message sends a TCP specific request from the Agent to Flare. At this time there are no options defined. At some time in the future the TCP interface may support network specific behavior such as having Flare generate Alerts for some conditions. All the Sub-tags other than the Request Id tag of this will be ignored in the initial version of this interface.

Tag = TAG_NET_OPTION_REQUEST
Sub Tag = TAG_NET_REQUET_ID

The TAG_NET_REQUEST_ID specifies a Request Id tag that Flare will return with the reply.

This option request can be sent at any time. In the future if the interface is changed and the old interface no longer supported, the option request maybe required to be the first message to allow Flare and the Agent to agree to use the new interface; an appropriate Tag will be defined with the new interface.

Other tags maybe added in the future as options are defined.

5.5 Option Reply

This message returns the result of an Option Request from Flare to the Agent. At this time the response will always be unsupported request. At some time in the future this request may be defined and the reply will return a status of zero to indicate successful request.

Tag = TAG_NET_OPTION_REPLY
Sub Tag = TAG_NET_REQUET_ID
Sub Tag = TAG_NET_OPTION_STATUS

The TAG_NET_REQUET_ID returns the Id of the request that generated this reply.

The TAG_NET_OPTIONS_STATUS returns a value of 1 indicating request not supported.

This message will also be sent by Flare to inform the Agent of problems with the connection. If Flare receives a connection and finds it can not service the connection, Flare will send an Option Reply

message with a status indicating why it will not service the connection. Thus if the Agent were to send an Option Request as the first request to Flare, a reply with a failure status will indicate why the connection is being closed.

The following values are defined for the status:

0 Request accepted

1 Request not supported (or tag in request not supported)

2 Connection not accepted, authorization failure

3 Connection not accepted, too many connections

6 Connection management

Flare will listen for TCP connections on port 2918. It will accept all connections. If Flare already has more than the implementation maximum (at least 8), it will send a TAG_NET_OPTION_REPLY with a status value of 3 indicating that the maximum number of connections has been exceeded and then close the connection. If Flare will not accept the connection because the agent's IP address is not in the list of allowed hosts, Flare will send a TAG_NET_OPTION_REPLY with a status value of 2.

When reading connections, Flare will implement a timeout. If Flare ever waits more than 100 seconds to finish reading a complete message, it will close the connection. The timeout does not start until the first byte of the message has been received; thus an idle connection can remain indefinitely. Flare will enable TCP keepalives on the connection to cleanup connection from hosts that are no longer accessible.

If Flare ever finds an invalid length on a Request Id, or SCSI CDB, it will assume the Agent and Flare are out of sync due to some problem and it will close the connection.

Appendix A: TAG LIST

TBD

End of Document